

Card fraud detection using machine learning

Class imbalance : 875 Fraud transactions out of total 118,621 (0.74%)

Approach : Feature engineering and extraction from raw data based on transaction time and amount

Raw data
Transaction Time
Account Number
MerchantId
mcc
MerchantCountry
Merchantzip
PoSEntryMode
TransactionAmount
AvailableCash

Feature engineering
and extraction



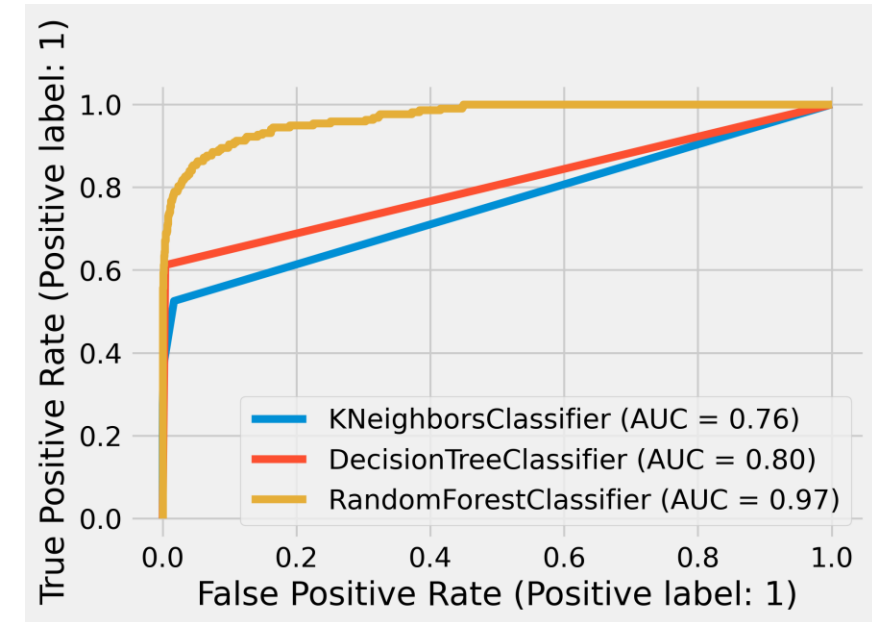
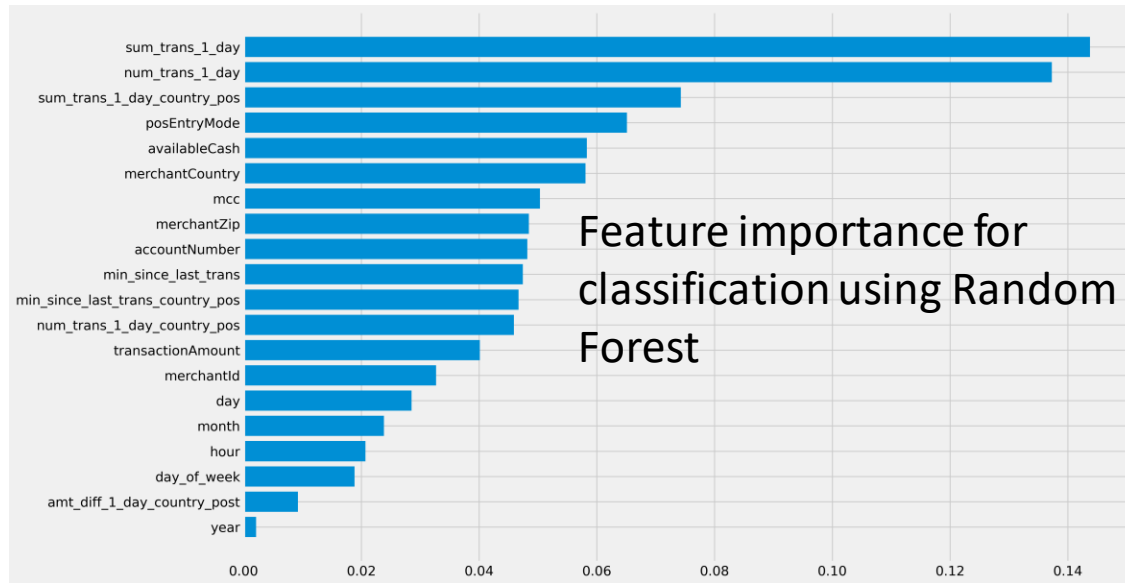
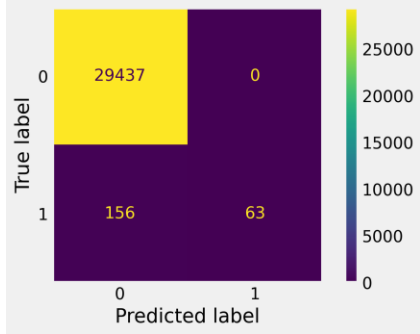
New Features based on time	
Feature	Explanation
Transaction year	Year of transaction
Transaction month	Month of transaction
Transaction day	Day of transaction
Transaction hour	Hour of transaction
Transaction day of the week	Day of the week of transaction
min_since_last_trans	Minutes passed since last transaction was made from an account
min_since_last_trans_country_pos	Minutes passed since last transaction was made from an account grouped according to country and pos type

New Features based on the transaction amount	
Feature	Explanation
amt_diff_1_day_country_pos	Difference in transaction amount in the last 24 h from an account grouped by country and pos type
num_trans_1_day	Number of transactions in the last 24 h from an account
sum_trans_1_day	Sum of the transaction amount in the last 24 h from an account
num_trans_1_day_country_pos	Number of transactions in the last 24 h from an account grouped according to the country and pos type
sum_trans_1_day_country_pos	Sum of the transaction amount in the last 24 h from an account grouped by country and pos type

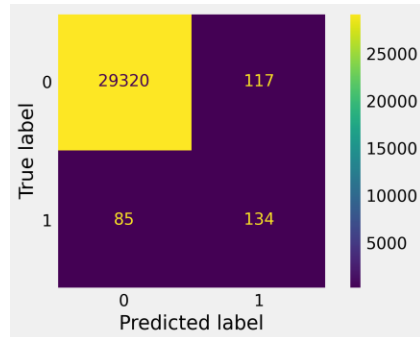
Features used for classification --> all new features + old features

Dataset split into 75% training and 25% test set in a stratified manner

Random Forest



Decision Tree



```

|--- num_trans_1_day <= 40.00
|   |--- sum_trans_1_day <= 1072.12
|       |--- merchantCountry <= 833.00
|           |--- posEntryMode <= 80.50
|               |--- posEntryMode <= 1.50
|                   |--- min_since_last_trans <= 29.50
|                       |--- accountNumber <= 5.50
|                           |--- sum_trans_1_day_country_pos <= 297.66
|                               |--- class: 0
|                               |--- sum_trans_1_day_country_pos > 297.66
|                                   |--- class: 1
|                                       |--- accountNumber > 5.50
    
```

Rules used to split at a node by Decision Tree classifier

Important observations

1. Aim: Minimum false negatives (fraud transactions predicted as not fraud)
2. DT classifier performs better than RF

$$FN_{DT} < FN_{RF}$$

$$TP_{DT} > TP_{RF}$$

3. Important features e.g. *number and sum of transaction in the last 24h* from an account are created features by feature engineering

Decision Trees achieves an accuracy of more than 100% than Random Forest in identifying fraudulent transactions